Photon Counting & the Statistics of Light

September 23, 2014

Jung Lin (Doris) Lee
dorislee@berkeley.edu
Group partners: Jennifer Ito, Manuel Silvia
Prof. James Graham, UGSI Heechan Yuk, Isaac Domagalski

Abstract

In our experimental setup, photons emmited by an LED light source is regestered by the PMT as a timestamp for every photon detection. By increasing the brightness of the light source, we found that photon arrival rates increase. Analysis on a single dataset shows that as the the number of events in an experiment (N) increases, the sample mean($\bar{x}$) converges to the theoretical mean(μ). The fluctuation around this value of μ falls off as with $1/\sqrt{N}$, so increasing N can results in more reliable experimental results. However, systematic effects can not be eliminated by such methods. We instead mitigate the afterpulsing effect by vetoing events that occurred immediately after another. The processed data results in a histogram that align very closely to exponential distribution. In addition, binned data for a fixed time interval also matches closely with the Poisson distribution. Finally, we can obtain the sample mean parameter for this experiment by fitting the histogram to the Poisson distribution.

1. Introduction

Large sky surveys such as the Sloan Digital Sky Survey(SDSS) and the Dark Energy Survey (DES) have surveyed large fractions of the sky providing detailed photometric and astrometric information for millions of sources. Much like how photons from an LED arrive at the PMT in our experiment, images are formed by individual photons that reaches the CCD. Therefore, some of the surveys' calibration pipeline relies on the knowledge of photon statistics, as they are governed by the same set of underlying noise and distribution. An understanding of how these systematic may affect the imaging data may be helpful in studying individual galaxy morphology, determining the transit of an exoplanet, or compiling brightness-limited catalogs for cosmological studies.

In this report, we present the methods used to conduct the photon counting experiment and the obtained statistical results. In section 2, I will detail the experimental method that my group used to take

the data. Then, section 3 introduces the various statistical measures that were used to describe the findings on a single dataset. Section 4 introduces the two probability density functions (PDFs) dthat are relevant to the histograms of these datasets.

2. Experiments

2.1. Methods

In the setup, the light from the LED is directed into a optical fiber into the photomultiplier tube (PMT) to minimize losses. A squawker and oscilloscope is connected to provide audio and visual feedback for a photon detection. The arrival of a photon is registered by a PMT, and data was acquired using the CoinPro software. The resulting data file is a series of timestamps corresponding to when the photon arrived. Since the CoinPro software is ran on a 32-bit machine, we can see the counts cycling through of time, with the extrema corresponding to the maximum and minimum 32-bit signed integer. Although

absolute timestamp is the recorded quantity, it is often more useful to consider the time elapsed between two consecutive events (`dt`) as plotted in Fig.1 in order to get a sense of the photon arrival rate for some period of waiting time.
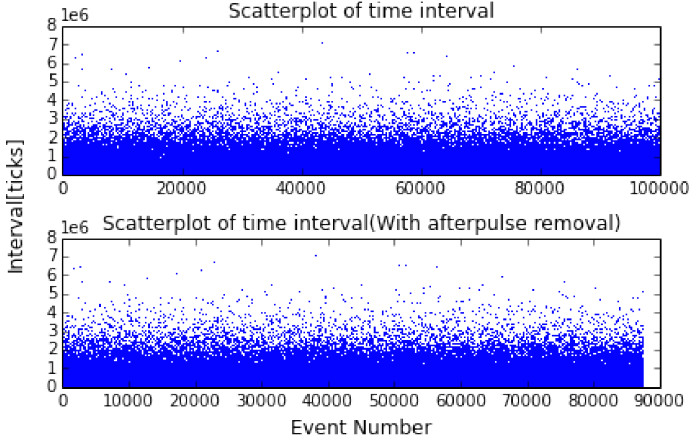

Figure 1: With the removal of after pulse events, we can see that there is a decline in the number of events. This makes sense because the afterpulse removal procedure cuts away all the interval that lasts 3000 clock ticks or less. As a result of this cut, we expect to see a clear boundary at y=3000, but, due to the large range of interval value that the y-axis spans over, this can not be seen. The data representation in Fig.3 and 6 better shows the the effect of gating afterpulses.

Sometimes noise pulses resembling Fig.2 can be observed after the original pulse. These events are caused by the ionization of residual gas molecules in the PMT's vacuum chamber. We eliminate this systematic effect by cutting away data with subsequent pulses lying within the time frame of $2.5\mu s$ after an original pulse as shown in Fig.1.

2.2. *Changing LED Brightness*

In order to take a set of data with different brightness level, we listened to ticks on the squawker and found that there was no detectable difference from the 3 o'clock off position to the 12 o'clock position of the LED brightness knob. Therefore, we started off at the 12 o'clock location and went to the max location (at around 90°), diving this region into 5 equal parts for the 6 brightness settings[1] of roughly 18°increments. Even though these measurements of

---

[1]Although it seems like only 5 data points are shown in the plot, this is merely an artifact of the choice of data representa-
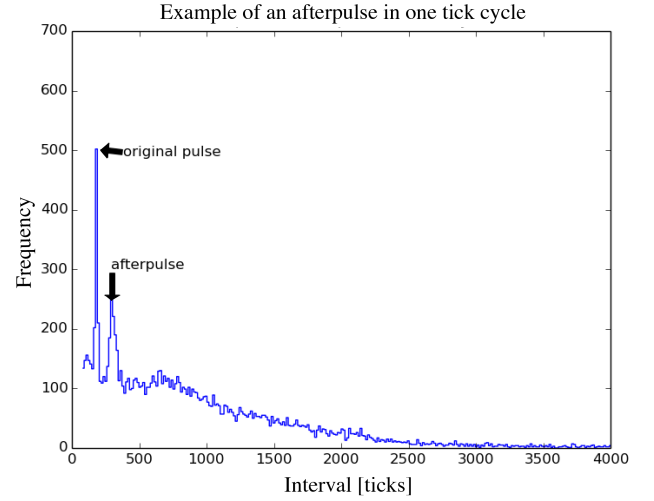

Figure 2: With a histogram of bin size = 500,000, the two peaks (original and afterpulse) as well as a gradual decline of ticks in the next 3000 clock ticks can be seen here.

Intensity[°]	Final timestamp in time series [ticks]
0	86447071309
18	68475846523
36	35127973096
54	7517634481
72	5787954870
90 (max)	5558407093

Table 1: Intensity is specified as the LED control knob position in degrees, as described earlier in Sec.2.2. Since we are taking a fixed number of photon arrival events, as the brightness increases, the faster photon arrival rate means that 10,000 event count finishes faster than it did for lower brightness; thus, accounting for the smaller final timestamp values.

the LED brightness were only estimated, it did not significantly affect the plot since the data points should fall on the $\sigma = \bar{x}$ line as the σ and μ of an exponential distribution are always equal.

3. Descriptive Statistics

3.1. *Simple Descriptions of Data*

Using the largest dataset we collected with the LED on maximum brightness, we obtain a sample mean of 561571 ticks and a standard deviation of

---

tion as circles as the first two data points lie so closely that they are almost indistinguishable. This is easily verified by using the point representation with `matplotlib's` ','setting
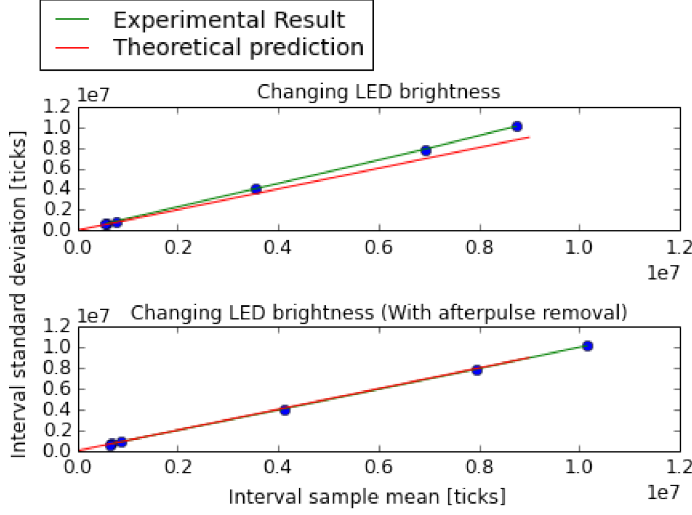
Figure 3: The slope of the line should be one since both the standard deviation and the sample mean of each dataset is equal to the single-parameter τ in the exponential distribution. We find that gating the afterpulses causes the $\bar{x}$ to get closer to μ.

637988 ticks on the whole dataset of 99833 events.[2] This is the most precise statistical quantity we can obtain using our largest dataset, since the more data that we compute with the closer the sample mean converges to the mean value as shown in Fig.4. Although higher precision measurement of $\bar{x}$ and σ can be obtained by taking a sample of larger N so that $\bar{x}$ approaches μ, but this does not ameliorate any systematic effect such as dark current or afterpulses.
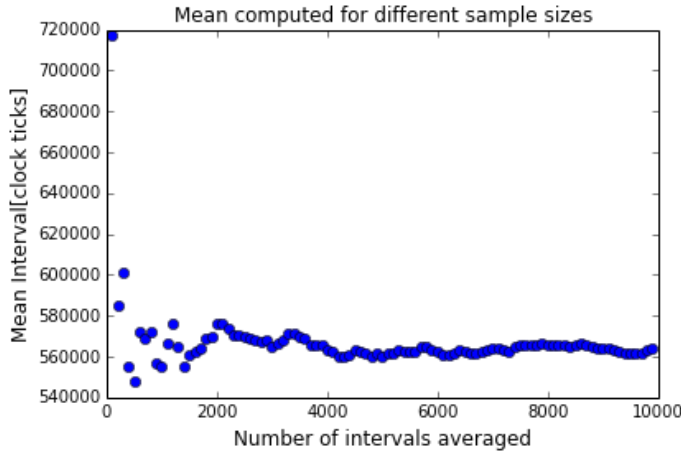

Figure 4: Large fluctuation in computed mean decreases as larger datasets are used to compute the mean.

3.2. Standard Deviation of the Mean (SDOM)

The SDOM tells us the uncertainty on the sample mean obtained from the given set of N datapoints. The $1/\sqrt{N}$ behavior shown in Fig.5 is important since it motivates an experiment with greater number of events, as it increases the confidence level in the experimental results.
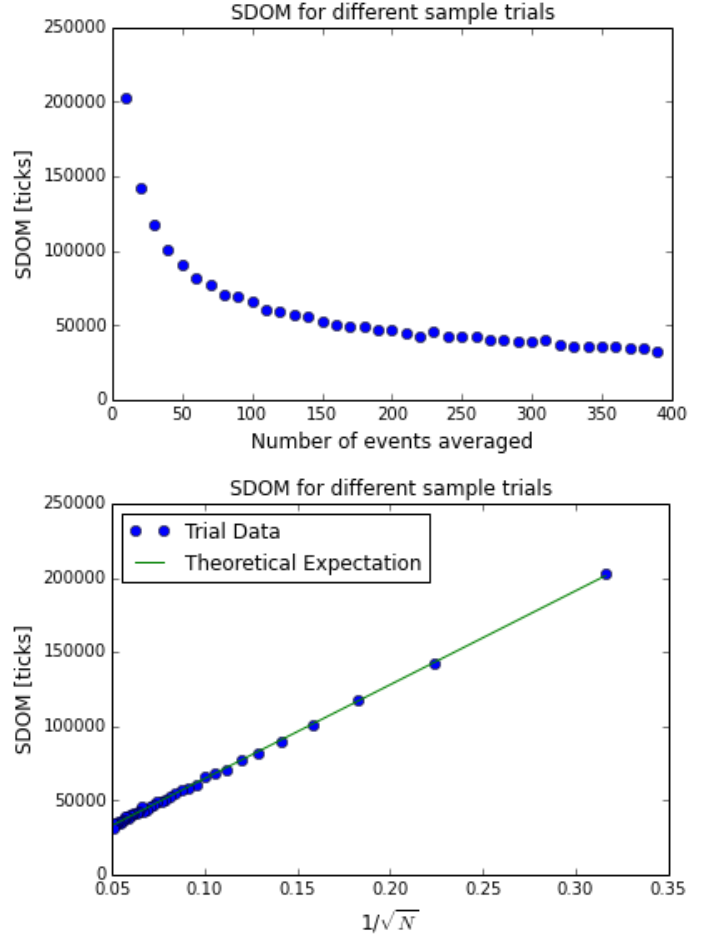

Figure 5: The top graph shows how SDOM declines as a function of the number of events. The resulting curve resembles an inverse-root relation. We test this hypothesis by transforming the x-axis into $1/\sqrt{N}$ and we indeed observe a linear relationship with the sample standard deviation as the slope.

4. Photon Statistics

Even though mean and standard deviation are useful descriptive quantities that characterizes the data, they are actually properties of the underlying distribution. Despite the inevitable loss of details from the binning procedure, histograms can often provide additional information about the distribution of data. In

this section, I will explain how different types of binning leads to the exponential and Poisson distribution and how they are related.

4.1. Exponential Distribution

Exponential distribution described by Eq.**??** governs the histogram of the elapsed time between photon arrival event (dt).

$$P(t, \tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \tag{1}$$

where Δt =time step-size and τ =mean interval.
A defining characteristic of the exponential distribution is that it is continuous"memorylessness": photon arrival rate at some moment in time,t, is not dependent on the photon arrival rate prior to time t. The green curve in Fig.6 is the plotted from the analytical solution shown in Eq.2. This result is obtained from integrating the exponential distribution with a $\Delta t = 1.42 \times 10^7$ and τ set as the mean of the dt dataset .

$$\int_{\tau}^{t+\Delta t} \frac{1}{\tau} e^{\frac{t}{\tau}} dt = -e^{\frac{t+\Delta t}{\tau}} + e^{\frac{t}{\tau}} \tag{2}$$

Since the exponential PDF is normalized to one, this result is scaled by the number of events to fit the histogram.
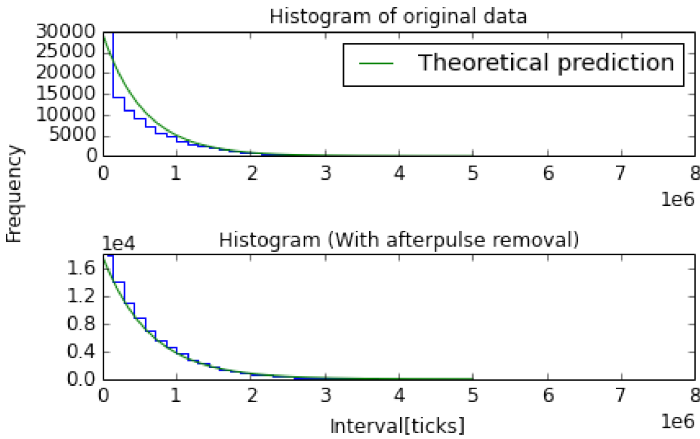

Figure 6: A histogram of bin size of 50 before and after gating afterpulses, with the scaled exponential PDF fit. Note that the histogram fits closer to the exponential distribution after the afterpulse removal.

4.2. Poisson Distribution

Although closely-related, the Poisson distribution is unlike the exponential distribution in that it is a discrete probability distribution that arises from measuring the number of events that occur within a fixed sampling interval. Both underlying process applies to rare event for very large N, but the experiment method is different. In this case, the experiment is the same, but we change the way the data is represented, as binned data for a fixed time interval.
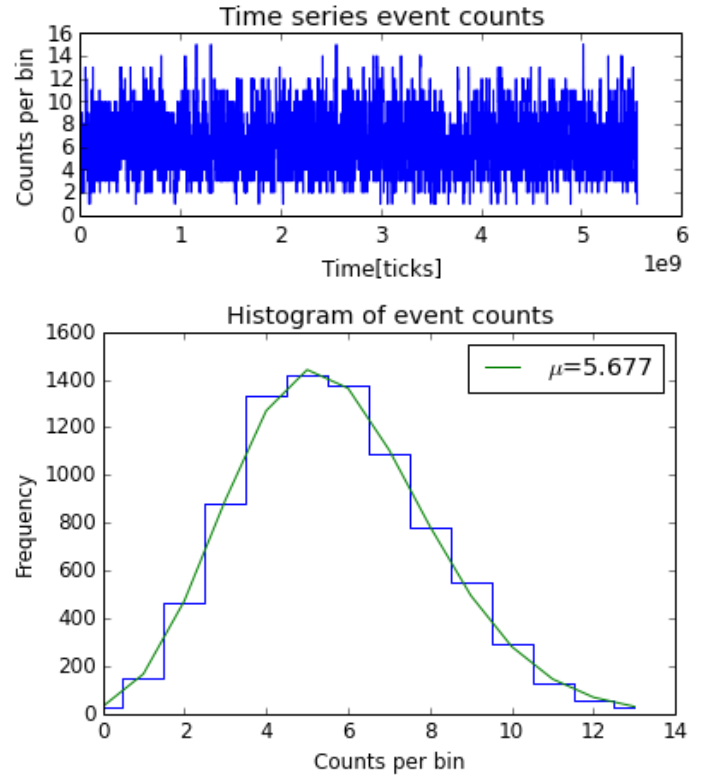

Figure 7: The bin counts are obtained by counting the number of photon arrival events that lie within an interval of 3.09 milliseconds. From the time series plot, we see that most time intervals counted 4 to 8 events. This is confirmed by using a Poisson curve fit on the histogram where the value of the fitting parameter is 5.677.

5. Conclusion

In this experiment, the apparatus is setup such that the PMT registers the photons emitted by the LED light source and the timestamp is recorded by the CoinPro program. We found that by changing the brightness of the light source, the photon arrival rate

4

increases, resulting in a shorter time series. By analyzing the interval between subsequent events, we observe interesting statistical properties that governs the datasets.

Since photon arrival is rare, random, and in the limit of large N, the mean interval histogram resemble closely to a exponential distribution. By binning the data with a fixed interval width and fitting a Poisson distribution over the resulting histogram, we can get the value of the $\bar{x}$ parameter. The Poissonian fluctuations in the photon arrival timing are yields an uncertainty of $\sqrt{N}$. Such fluctuations are random and independent of any systematic error. Therefore, we can not use the approach of increasing N to mitigate the afterpulse effect. Instead, we vetoed spurious ion events that followed immediately after a real photon emission and obtained more accurate results.

As we only attempt to mitigate the effect of one type of systematic error, one possible extensions to this project may be to investigate the effect of other types of systematic error such as dark counts and read noise on the quality of the data. This effect can be quantified by measuring the signal to noise ratio and comparisons among different types/manufacturers of PMTs.

6. References

Pickett, Siobhan, Sean Carriedo, and Chang Wang. "Determining the Signal-to-Noise Ratio and Optimal Photomultiplier Gain Setting in the GenePix 4000B." *Genomics Application Note* (2001)